Applied Microeconometrics Problem Set 2

Tanya Rajan

November 10, 2020

Question 1 Let $\mathcal{H}$ denote the set of all proper distribution functions for a scalar random variable. Suppose that we observe a non-negative random variable Y with distribution $G \in \mathcal{G}$ where $\mathcal{G}$ is the subset of $\mathcal{H}$ such that $\mathbb{P}_G[Y \geq 0] = 1$. Assume that Y is determined by

$$Y = \max\{U, 0\}$$

where U is an unobserved random variable with distribution function F. Assume that $F \in \mathcal{F}$, where $\mathcal{F}$ is some subset of $\mathcal{H}$ specified below. Notice that any $F \in \mathcal{F}$ implies a distribution for Y called G_F, defined as

$$G_F(y) \equiv \mathbb{P}_F[\max\{U, 0\} \leq y]$$

Let $\mathcal{F}^*(G)$ denote the sharp identified set for F, that is

$$\mathcal{F}^*(G) = \{F \in \mathcal{F} : G_F(y) = G(y) \quad \text{for all } y \in \mathbb{R}\}$$

Consider the target parameter $\pi : \mathcal{F} \to \mathbb{R}$ defined as $\pi(F) = \mathbf{E}_F[U]$. Let $\Pi^\star(G)$ denote the sharp identified set for π.

In this problem we are asked to show whether the model is falsifiable under different restrictions on the unobserved distribution of F. Falsifiability means that there is a function $\tau : \mathcal{F} \to \mathbb{R}$ s.t. $\tau(G) = 1 \implies \mathcal{F}^*(G) = \emptyset \implies \Pi^*(G) = \emptyset$ and $\tau(G) = 1$ for at least one $G \in \mathcal{G}$. So given the restrictions on $\mathcal{F}$, we want to see whether there is some G with $\mathbb{P}_G(Y \geq 0) = 1$ for which $G_F(y) \neq G(y)$ for some $y \in \mathbb{R}$ and for all $F \in \mathcal{F}$.

(a) Suppose that $\mathcal{F} = \mathcal{H}$. Determine $\Pi^*(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?

$$\Pi^*(G) = \{\mathbb{E}_F[U] : F \in \mathcal{H}, G_F(y) = G(y) \quad \forall y \in \mathbb{R}\}$$

This problem puts no restrictions on the set of hypothetical distributions $\mathcal{F}$. $F \in \mathcal{F}$ can be any proper distribution function for a scalar random variable. We are told that $\mathcal{G} \subseteq \mathcal{H} = \mathcal{F}$, so for any $G \in \mathcal{G}$, we can trivially take $F = G \in \mathcal{F}$. Since G is constrained to be a distribution over nonnegative support, $F(y) = \mathbb{P}_F(\max\{U, 0\} \leq y) = G_F(y)$. As a result, we always have $G_F(y) = G(y)$ regardless of which $G \in \mathcal{G}$ we pick. So the model is not falsifiable since $\mathcal{F}^*(G)$ (and as a result $\Pi^*(G)$) is never empty.

(b) Suppose that $\mathcal{F} = \{F \in \mathcal{H} : F(-1) = 0 \text{ and } F(2) = 1\}$. Determine $\Pi^\star(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?

$$\Pi^*(G) = \{\mathbb{E}_F[U] : F \in \mathcal{H}, F(-1) = 0, F(2) = 1, G_F(y) = G(y) \quad \forall y \in \mathbb{R}\}$$

Part (b) restricts $F \in \mathcal{F}$ such that $\text{supp}(F) = [-1, 2]$. This means that $G_F(y) = \mathbb{P}(\max\{U, 0\} \leq y)$ has a support of $[0, 2]$. To find a $G \in \mathcal{G}$ for which $\mathcal{F}^*(G)$ is empty, we simply have to find a distribution such that

$[0, 2] \subset \text{supp}(G)$. In this case, no $F \in \mathcal{F}$ s.t. $G_F = G$ would be able to satisfy $G_F = G$, so $\mathcal{F}^*(G)$ and $\Pi^*(G)$ are empty. Since $G \in \mathcal{G}$ is only defined on positive support, we can formally state the falsifiability test as $\tau(G) = \mathbb{1}\{G(2) < 1\}$. So we have:

$$\Pi^*(G) = \{\mathbb{E}_G[U] : G \in \mathcal{G}, G(2) < 1\}$$

(c) Suppose that $\mathcal{F} = \{F \in \mathcal{H} : F(0) = 1/2\}$. Determine $\Pi^\star(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?

$$\Pi^*(G) = \left\{\mathbb{E}_F[U] : F \in \mathcal{H}, F(0) = \frac{1}{2}, G_F(y) = G(y) \quad \forall y \in \mathbb{R}\right\}$$

We are told that F must be a distribution where half the observations are less than or equal to 0. Then we can say that

$$F(0) = \mathbb{P}_F(U \leq 0) = \mathbb{P}_F(\max\{U, 0\} \leq 0) = G(0) = \frac{1}{2}$$

So a test for falsifiability of this model is $\tau(G) = \mathbb{1}\left\{G(0) \neq \frac{1}{2}\right\}$. The model is falsifiable by this standard. So we have:

$$\Pi^*(G) = \{\mathbb{E}_G[U] : G \in \mathcal{G}, G(0) = \frac{1}{2}\}$$

.

(d) Suppose that $\mathcal{F} = \{F \in \mathcal{H} : \mathbb{E}_F[U] = 0\}$. Determine $\Pi^\star(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?

$$\Pi^*(G) = \{\mathbb{E}_F[U] : F \in \mathcal{H}, \mathbb{E}_F[U] = 0, G_F(y) = G(y) \quad \forall y \in \mathbb{R}\}$$

Part (d) restricts F to be a distribution with an expected value of 0. We can write:

$$\mathbb{E}_F[U] = 0 \implies \mathbb{E}_{G_F}[Y] = \mathbb{E}_F[\max\{U, 0\}] \geq 0$$

In order for $\mathbb{E}_F[U] = 0$, we need either that

1. There is a point mass at the point $u = 0$
2. F must have some negative and some positive support. In other words, $\text{supp}(F) \cap (-\infty, 0) \neq \emptyset$ and $\text{supp}(F) \cap (0, \infty) \neq \emptyset$.

In case (1), $\mathbb{E}_F[\max\{U, 0\}] = \mathbb{E}_{G_F}[Y] = 0$ and G_F also has a point mass at 0. In case (2), $\mathbb{E}_{G_F}[Y] > 0$.

Additionally, since $G \in \mathcal{G}$ is restricted to be a proper distribution on nonnegative support, it must be that $\mathbb{E}_G[Y] \geq 0$. If $\mathbb{E}_G[Y] = 0$, it must have a point mass at 0, so we can trivially set $F = G$ to satisfy $G_F(y) = G(y)$. However, if $\mathbb{E}_G[Y] > 0$, and $G(0) = 0$, it is impossible to find a distribution F that can generate $G(y) = G_F(y)$. This is because the probability at all the negative points of support of F are shifted to 0 by G_F due to the max operator. As a result, the point 0 will have some non-zero probability s.t. $G_F(0) \neq 0$. So the falsifiability test is: $\tau(G) = \mathbb{1}\{G(0) = 0 \cap \mathbb{E}_G[Y] > 0\}$. Then we have:

$$\Pi^*(G) = \{\mathbb{E}_G[U] : G \in \mathcal{G}, G(0) = 0 \cap \mathbb{E}_G[Y] > 0\}$$

.

(e) Now change the target parameter to $\pi(F) \equiv \text{med}_F(U) \equiv \inf\left\{u : F(u) \geq \frac{1}{2}\right\}$ (the median of U when it's distributed like F). Suppose that $\mathcal{F} = \mathcal{H}$. Determine $\Pi^\star(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?

$$\Pi^*(G) = \{\text{med}_F(U) : F \in \mathcal{H}, G_F(y) = G(y) \quad \forall y \in \mathbb{R}\}$$

This case is the same as part (a). Even though we have a different target parameter, there are not enough restrictions on the possible distributions $\mathcal{F}$ to allow us to falsify the model. As in part (a), we can always trivially take $F = G$ to make the equality $G_F(y) = G(y)$ hold. The model is not falsifiable.

Question 2 Consider the simple instrumental variables model

$$Y = \alpha X + U$$

where X is scalar. Suppose that the instrument, Z is also scalar, and that there is homoskedasticity with respect to Z in both the reduced form and first stages. (This is the case used to discuss weak instruments in the supplemental notes.) Let $\hat{F}$ denote the sample first stage $F-$ statistic. Determine the asymptotic bias of $\hat{F}$ as an estimator of the concentration parameter $\left(\mu^{2n}\right)$ under weak instrument asymptotics. Construct an alternative estimator that is asymptotically unbiased. Hint: No need to reinvent what I did in the supplemental notes. You may use any of the derivations there. Constructing the alternative estimator should be easy.

From the supplemental notes, we have that $\hat{F}$ is defined as:

$$\hat{F} \equiv n\frac{\hat{\pi}^2}{\hat{\sigma}_\pi^2} = \left(\frac{\sqrt{n}\hat{\pi}}{\hat{\sigma}_\pi}\right)^2$$

We are told that $\hat{\sigma}_\pi$ is a consistent estimator for σ_π. Further, we are given that (by WLLN, CMT, Slutsky's) $\sqrt{n}\hat{\pi}$ converges in distribution to:

$$\sqrt{n}\hat{\pi} \equiv \sqrt{n}\pi_n + \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i V_i}{\left(\frac{1}{n}\sum_{i=1}^n Z_i^2\right)} \xrightarrow{d} \pi + \frac{R_{\text{fs}}}{\mathbb{E}\left[Z^2\right]}$$

Since $\sqrt{n}\hat{\pi} \xrightarrow{d} \pi + \frac{R_{\text{fs}}}{\mathbb{E}[Z^2]}$ and $\hat{\sigma}_\pi \xrightarrow{p} \sigma_\pi$, we can use CMT and Slutsky's Lemma (when $\sigma_\pi \neq 0$) to say that:

$$\begin{aligned}
\hat{F} = \left(\frac{\sqrt{n}\hat{\pi}}{\hat{\sigma}_\pi}\right)^2 &\xrightarrow{d} \left(\frac{\pi + \frac{R_{fs}}{\mathbb{E}[z^2]}}{\sigma_\pi}\right)^2 \\
&= \frac{\pi^2 + 2\pi\frac{R_{fs}}{E[Z^2]} + \frac{R_{fs}^2}{\mathbb{E}^2[Z^2]}}{\sigma_\pi^2} \\
&= \frac{\pi^2\mathbb{E}\left[Z^2\right]}{\sigma_V^2} + \frac{2\pi R_{fs}}{\sigma_V^2} + \frac{R_{fs}^2}{\sigma_V^2} \\
&= \mu^2 + \frac{2\pi R_{fs}}{\sigma_V^2} + \frac{R_{fs}^2}{\sigma_V^2}
\end{aligned}$$

Note the second to last line follows from $\sigma_\pi^2 = \frac{\sigma_V^2}{\mathbb{E}[Z^2]}$. So the asymptotic bias for $\hat{F}$ as an estimator for μ^2 is:

$$\begin{aligned}
|\operatorname{ABias}(\hat{F})| = |\lim_{n\to\infty}\mathbb{E}[\mu^2 - \hat{F}]| &= \left|\lim_{n\to\infty}\mathbb{E}\left[\mu^2 - \mu^2 - \frac{2\pi R_{fs}}{\sigma_V^2} - \frac{R_{fs}^2}{\sigma_V^2}\right]\right| \\
&= \left|\lim_{n\to\infty}\mathbb{E}\left[\frac{2\pi R_{fs}}{\sigma_V^2} + \frac{R_{fs}^2}{\sigma_V^2}\right]\right|
\end{aligned}$$

An alternative unbiased estimator would counteract the extra term from $\sqrt{n}\hat{\pi}$. Define the following:

$$\hat{F}^* = \left(\frac{\sqrt{n}\hat{\pi} - \frac{\frac{1}{n}\sum_{i=1}^n Z_i\hat{V}_i}{\frac{1}{n}\sum_{i=1}^n Z_i^2}}{\hat{\sigma}_\pi}\right)^2$$

where $\hat{V}_i = X_i - \hat{\pi} Z$. Since $\hat{\pi}$ is a consistent estimator for π, by WLLN and CMT $\frac{1}{n}\sum_{i=1}^n Z_i \hat{V}_i \xrightarrow{p} \mathbb{E}[Z_i V_i] = 0$. We know by WLLN that $\frac{1}{n}\sum_{i=1}^n Z_i^2 \xrightarrow{p} \mathbb{E}[Z^2]$. Then we can argue as follows:

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i \hat{V}_i \xrightarrow{d} R_{fs} \qquad \text{by definition of } R_{fs}$$

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n Z_i \hat{V}_i}{\frac{1}{n}\sum_{i=1}^n Z_i^2} \xrightarrow{d} \frac{R_{fs}}{\mathbb{E}[Z^2]} \qquad \text{by Slutsky's Lemma}$$

$$\hat{F}^* \xrightarrow{d} \left(\frac{\pi + \frac{R_{fs}}{\mathbb{E}[Z^2]} - \frac{R_{fs}}{\mathbb{E}[Z^2]}}{\sigma_\pi}\right)^2 = \left(\frac{\pi \mathbb{E}[Z^2]}{\sigma_V}\right)^2 = \mu^2 \qquad \text{by CMT, Slutsky's}$$

Question 3 Consider the binary treatment potential outcomes model $Y = DY(1) + (1-D)Y(0)$. Suppose that we have a binary treatment Z that is independent of $(Y(0), Y(1), D(0), D(1))$ where $D = ZD(1) + (1-Z)D(0)$. Maintain the Imbens and Angrist (1994) monotonicity condition that $D(1) \geq D(0)$, and assume that $\mathbb{P}[D = 1 \mid Z = 1] - \mathbb{P}[D = 1 \mid Z = 0] > 0$ Let $G \in \{c, a, n\}$ denote the types complier, always-taker and never-taker, i.e. $G = c$. Let $F_d(y \mid g)$ denote the distribution function for $Y(d)$ conditional on $G = g$

(a) Show that $\mathrm{F}_1(y \mid G = a)$ is point identified for any y .

$$\begin{aligned}
\mathbb{P}(Y(1) \leq y \mid G = a) &= \mathbb{P}(Y(1) \leq y \mid D(0) = D(1) = 1) \\
&= \mathbb{P}(Y \leq y \mid D(0) = D(1) = 1) \\
&= \mathbb{P}(Y \leq y \mid D(0) = 1) && \text{by monotonicity} \\
&= \mathbb{P}(Y \leq y \mid 0 = 1, Z = 0)
\end{aligned}$$

This is observed in the data and has a specific value for each possible y.

(b) Show that $\mathrm{F}_0(y \mid G = n)$ is point identified for any y .

$$\begin{aligned}
\mathbb{P}(Y(0) \leq y \mid G = n) &= \mathbb{P}(Y(0) \leq y \mid D(0) = D(1) = 0) \\
&= \mathbb{P}(Y \leq y \mid D(0) = D(1) = 0) \\
&= \mathbb{P}(Y \leq y \mid D(1) = 0) && \text{by monotonicity} \\
&= \mathbb{P}(y \leq y \mid D = 0, Z = 1)
\end{aligned}$$

This is observed in the data and has a specific value for each possible y.

(c) Show that both $\mathrm{F}_1(y \mid G = c)$ and $F_0(y \mid G = c)$ are point identified for any y .

We have already shown that $F_1(y|G = a)$ and $F_0(y|G = n)$ are point identified. We will use this in the derivations for this part. Additionally, the following probabilities are observed in the data:

$$\begin{aligned}
p_a &= \mathbb{P}(D = 1|Z = 0) = \frac{\mathbb{P}(D = 1, Z = 0)}{\mathbb{P}(Z = 0)} \\
p_n &= \mathbb{P}(D = 0|Z = 1) = \frac{\mathbb{P}(D = 0, Z = 1)}{\mathbb{P}(Z = 1)} \\
p_c &= 1 - p_a - p_n
\end{aligned}$$

The final line is true because the monotonicity condition assumes away defiers, so there are only compliers, always-takers, and never-takers. Now let us examine $F_1(y|c)$:

$$\begin{aligned}
\mathbb{P}(Y \leq y \mid D = 1, Z = 1) &= \mathbb{P}(Y(1) \leq y \mid D(1) = 1, D(0) = 0, Z = 1)\mathbb{P}(D(1) = 1, D(0) = 0 \mid Z = 1, D = 1) \\
&\quad + \mathbb{P}(Y(1) \leq y \mid D(1) = 1, D(0) = 1, Z = 1)\mathbb{P}(D(1) = 1, D(0) = 1 \mid Z = 1, D = 1)
\end{aligned}$$

Note that the proportion $\mathbb{P}(Z=1, D=1)$ is the fraction of compliers plus the fraction of compliers. Then using the definition of a conditional probability, we get that $\mathbb{P}(D(1)=1, D(0)=0 \mid Z=1, D=1) = \frac{p_c}{p_c+p_a}$ and $\mathbb{P}(D(1)=1, D(0)=1 \mid Z=1, D=1) = \frac{p_a}{p_c+p_a}$. Rewriting the equation above, we have:

$$\begin{aligned}
\mathbb{P}(Y \leq y \mid D=1, Z=1) &= F_1(y \mid G=c)\frac{pc}{p_c+p_a} + F_1(y \mid G=a)\frac{p_a}{p_c+p_a} \\
\implies F_1(y|G=c) &= \left(\mathbb{P}(Y \leq y \mid D=1, Z=1) - F_1(y \mid G=a)\frac{p_a}{p_c+p_a}\right)\frac{p_c+p_a}{p_c} \\
&= \mathbb{P}(Y \leq y \mid D=1, Z=1)\frac{p_c+p_a}{p_c} - F_1(y \mid G=a)\frac{p_a}{p_c}
\end{aligned}$$

All of the terms in this equation are observed in the data and can be point identified for a given y. Following a similar process for $F_0(y)$, we get:

$$\begin{aligned}
\mathbb{P}(Y \leq y|D=0, Z=0) &= F_0(y|G=c)\frac{p_c}{p_c+p_n} + F_0(y|G=n)\frac{p_n}{p_c+p_n} \\
\implies F_0(y|G=c) &= \mathbb{P}(Y \leq y|D=0, Z=0)\frac{p_c+p_n}{p_c} - F_0(y|G=n)\frac{p_n}{p_c}
\end{aligned}$$

Again, all of these terms can be observed in the data so $F_0(y|G=c)$ is point identified for a given y.

> (d) Suppose that both $Y(0)$ and $Y(1)$ are continuously distributed scalar random variables with support over the entirety of the real line. Let F_1 and F_0 denote the unconditional distribution functions of $Y(1)$ and $Y(0)$. Define the random variables $U(0) \equiv F_0(Y(0))$ and $U(1) \equiv F_1(Y(1))$. Assume that $U(0) = U(1) = U$ with probability 1 Show that $F_0(y)$ and $F_1(y)$ are point identified for any y

This problem gives us a rank invariance assumption. The functions $U(0)$ and $U(1)$ are probability integral transforms of $Y(0)$ and $Y(1)$, respectively. Setting these two equal to each other is the same as saying that a person's rank in the distribution of outcomes is the same for both $Y(0)$ and $Y(1)$. Now we can turn to $F_0(y)$:

$$\begin{aligned}
F_0(y) &= \mathbb{P}(Y(0) \leq y) \\
&= \mathbb{P}(Y(0) \leq y|G=n)p_n + \mathbb{P}(Y(0) \leq y \mid G=c)p_c + \mathbb{P}(Y(0) \leq y \mid G=a)p_a
\end{aligned}$$

The only quantity in this equation that we have not previously identified is $\mathbb{P}(Y(0) \leq y \mid G=a)$. Considering only this term, we can write

$$\begin{aligned}
\mathbb{P}(Y(0) \leq y \mid G=a) &= \mathbb{P}\left(F_0(Y(0)) \leq F_0(y) \mid G=a\right) \\
&= \mathbb{P}\left(F_1(Y(1)) \leq F_0(y) \mid G=a\right) \\
&= \mathbb{P}(Y(1) \leq F_0(y) \mid G=a)
\end{aligned}$$

Appealing to rank invariance, we can use $F_0(y|G=c)$ to calculate what the value $F_0(y)$ is across the points of support of $Y(1)-Y(0)$. By similar arguments, for $F_1(y)$, the only term we have not identified is $\mathbb{P}(Y(1) \leq y|G=n)$. Using the same steps as above, we can show that this quantity is equal to $\mathbb{P}(Y(0) \leq F_1(y)|G=n)$.

> (e) Maintain the assumptions of the previous part. Let G denote the unconditional distribution function of $Y(1) - Y(0)$. Show that $G(y)$ is point identified for any y.

We want to find

$$G(y) = \mathbb{P}(Y(1) - Y(0) \leq y)$$

From the maintained assumptions, we can say:

$$\begin{aligned}
F_0(Y(0)) = U(0) = U(1) = F_1(Y(1)) \\
\implies Y(0) = F_0^{-1}\left(F_1(Y(1))\right) \\
\implies Y(1) = F_1^{-1}\left(F_0(Y(0))\right)
\end{aligned}$$

Note we can recover F_1, F_0 and their inverses by the arguments in part (d). Turning to the object of interest in this question:

$$G(y) = \mathbb{P}\left(Y(1) - F_0^{-1}\left(F_1(Y(1))\right) \leq y\right) = \mathbb{P}(F_1^{-1}(F_0(Y(0))) - Y(0) \leq y)$$

All of the quantities in the equation above are identified as shown in parts (a) - (d). So $G(y)$ is point-identified.

Question 4 This problem is about the paper "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" by Angrist and Evans (1998, The American Economic Review). The authors define their endogenous variable as a binary indicator for whether a family has exactly two or more than two children. They use two instruments: the same-sex instrument and the twin birth instrument, both separately and together.

(a) Discuss the interpretation and credibility of the monotonicity condition when using the same-sex instrument separately.

Here the monotonicity assumption requires that having two children of the same sex makes all individuals more (or less) likely to have a third child. This assumption is violated if same-sex birth induces some parents to stop having children while inducing others to have another child (i.e. the instrument shifts the treatment take-up in different directions for different individuals). The credibility of this assumption seems suspect. For example, if there is a societal preference for having male children as well as a secondary preference for having a two-child family, mothers who give birth to two boys may stop having children while mothers who give birth to two girls may continue having children. Since the instrument only looks at "same sex births", these two channels would be grouped together and could potentially cause the monotonicity assumption to fail.

The authors try to justify the monotonicity assumption by presenting descriptive evidence from Census and CPS data that women with two children of the same sex are more likely to have another child. In Table 3, they list the fraction of the sample who had another child conditional of the sex mix of the first two children. They show that those who have children of the same sex were more likely to have an additional child. The table also seems to suggest that mothers who have two girls are more likely to try for a third child than mothers who have two boys. However this evidence doesn't necessarily prove that the monotonicity assumption holds across all individuals since the fractions they report are averages.

(b) Discuss the interpretation and credibility of the monotonicity condition when using the twin birth instrument separately.

The monotonicity condition for the twin birth instrument requires that having a twin birth makes all individuals more (or less) likely to have a third child. This is fairly credible because having twins is an unanticipated shock that leaves parents with one more child than they planned to have. So "defiers" (who move in the opposite way from the compliers when faced with the instrument) are ruled out by construction.

(c) Discuss the interpretation and credibility of the monotonicity condition when using the same-sex and twin birth instruments together.

When combining these two instruments, we now have a vector $z = (z_{ss}, z_t)$ where z_{ss} is the same-sex instrument and z_t is the twin instrument. There are four possible values for this instrument: $\mathcal{Z} = \{(0,0),(0,1),(1,0),(1,1)\}$. For monotonicity to hold, we must have that for all $z, z' \in \mathcal{Z}$, $D_i(z) \geq D_i(z'), \forall i \in \mathcal{I}$ or $D_i(z) \leq D_i(z'), \forall i \in \mathcal{I}$. Let us take the vectors $z = (0,1)$, representing opposite-sex twin births and $z' = (1,0)$ representing same-sex non-twin births. It is unclear which of these vectors makes a mother more likely to have an extra child since these counterfactuals are hard to compare. In other words, we cannot guarantee that $D_i(z) \geq D_i(z')$ or $D_i(z) \leq D_i(z')$ for all individuals. The monotonicity condition for using both instruments together does not seem too credible.

Question 5 Consider an instrumental variables model with a binary treatment $D \in \{0,1\}$ and binary instrument $Z \in \{0,1\}$. The usual potential outcomes are $Y(0)$ and $Y(1)$ and the usual potential treatment choices are $D(0)$ and $D(1)$. Let $X \in \{1,\dots,K\}$ be a vector of covariates that has a discrete distribution with K points of support. Let β_{tsls} denote the population coefficient on D in the following two stage least squares specification:

$$\text{reduced form: } Y \text{ on } Z \text{ and } \{\mathbb{1}[X=k]\}_{k=1}^{K}$$
$$\text{first stage:} D \text{ on } Z \text{ and } \{\mathbb{1}[X=k]\}_{k=1}^{K}$$

and assume that β_{tsls} exists. Note that, relative to the case discussed in the slides and notes, the first stage in this specification does not contain interactions between Z and X. Maintain the usual exogeneity condition that Z is independent of $(Y(0), Y(1), D(0), D(1))$ conditional on X. Assume that $\mathbb{P}[D(1) \geq D(0) \mid X = k] = 1$ for all $k = 1,\dots,K$.

(a) Explain how the monotonicity condition given here differs from the one discussed in the slides.

The monotonicity condition in the slides conditioned on values of X but allowed the monotonicity to work in either direction. Stated more formally, the slides said

$$\mathbb{P}[D(1) \geq D(0) \text{ or } D(1) \leq D(0) \mid X] = 1$$

Here, however, we are saying that only one of the possibilities must be true for all values of X. The problem restricts us to the case where $\mathbb{P}[D(1) \geq D(0) \mid X = k] = 1$ for all $k = 1,\dots,K$.

(b) Show that

$$\beta_{\text{tsls}} = \mathbb{E}\left[\frac{\text{Cov}(D, Z \mid X)}{\mathbb{E}[\text{Cov}(D, Z \mid X)]}\mathbb{E}[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0, X]\right]$$

where $p(z,x) \equiv \mathbb{P}[D = 1 \mid Z = z, X = x]$ is the propensity score.

First note that we can rewrite the equation above as:

$$\beta_{\text{TSLS}} = \mathbb{E}\left[\frac{\text{Cov}(D, Z \mid X)}{\mathbb{E}[\text{Cov}(D, Z \mid X)]} LATE(X)\right]$$

So we want to show that the TSLS estimand is a weighted average of the LATE(X). First let us set up our two stage regression framework.

$$Y = \beta_0 + \beta_{TSLS} D + \beta_X + u$$
$$D = \pi_0 + \pi Z + \pi_X + v$$

where u, v are mean-zero error terms, $\beta_X = \sum_{k=1}^{K} \mathbb{1}\{X = x_k\}\beta_{x_k}$ and $\pi_X = \sum_{k=1}^{K} \mathbb{1}\{X = x_k\}\pi_{x_k}$. This setup is essentially running both stages of the 2SLS including dummies for each value of $X \in \{1,\dots,K\}$. We can write:

$$\beta_{TSLS} = \frac{\text{cov}(\tilde{Y}, \tilde{Z})}{\text{cov}(\tilde{D}, \tilde{Z})}$$

where all variables with tildes are of the form $\tilde{Y} = Y - \mathrm{BLP}\left(Y \mid \beta_X\right)$. Note that because we are conditioning on β_X, we have a fully saturated regression in X, which means that $BLP(Y|\beta_X) = \mathbb{E}[Y|X]$. So we can write that

$$\begin{aligned}
\tilde{Y} &= Y - \mathbb{E}[Y|X] \\
\tilde{Z} &= Z - \mathbb{E}[Z|X] \\
\tilde{D} &= D - \mathbb{E}[D|X]
\end{aligned}$$

Now let us go back to the β_{TSLS} expression above and consider the numerator separately.

$$\begin{aligned}
\operatorname{cov}(\tilde{Y}, \tilde{Z}) &= \operatorname{cov}(Y - \mathbb{E}[Y \mid X], Z - \mathbb{E}[Z \mid X]) \\
&= \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y \mid X])(Z - \mathbb{E}[Z \mid X])/X]] \\
&= \mathbb{E}[\mathbb{E}[YZ \mid X] - \mathbb{E}[Y \mid X]\mathbb{E}[Z \mid X]] \\
&= \mathbb{E}[\operatorname{cov}(Y, Z \mid X)]
\end{aligned}$$

Similarly, we can derive $\operatorname{cov}(\tilde{D}, \tilde{Z}) = \mathbb{E}[\operatorname{cov}(D, Z|X)]$. Now using the formula for the LATE estimand, we get:

$$\begin{aligned}
\mathrm{LATE}(X) &= \frac{\operatorname{cov}(Y, Z \mid X)}{\operatorname{cov}(D, Z \mid X)} \\
\implies \operatorname{cov}(\tilde{Y}, \tilde{Z}) &= \mathbb{E}[\mathrm{LATE}(X) \operatorname{cov}(D, Z \mid X)]
\end{aligned}$$

Putting this all together, we can rewrite β_{TSLS} as:

$$\begin{aligned}
\beta_{TSLS} &= \frac{\mathbb{E}[\mathrm{LATE}(X) \operatorname{cov}(D, Z \mid X)]}{\mathbb{E}[\operatorname{cov}(D, Z \mid X)]} \\
&= \mathbb{E}\left[\frac{\mathrm{LATE}(X) \operatorname{cov}(D, Z \mid X)}{\mathbb{E}[\operatorname{cov}(D, Z \mid X)]}\right]
\end{aligned}$$

So we have the desired result.

(c) Explain what this result shows and provide some intuition.

As discussed briefly above, the TSLS estimand is a weighted average of LATE(X). The weights are determined by how much the instrument and the treatment co-vary given X. Subgroups $X = x$ for which D and Z have a high covariance after controlling for X will receive a higher weight. So groups that have more people induced into treatment by the instrument (i.e. more compliers) will be weighted higher.

Question 6 Consider the following data generating process:

$$Y = X + U$$
$$X = .3Z_1 + V$$

where (U, V) are jointly normal, mean zero, with variance matrix given by

$$\begin{bmatrix} .25 & .20 \\ .20 & .25 \end{bmatrix}$$

Let $Z_1, Z_2, \ldots, Z_{20}$ be independent standard normals. Run a Monte Carlo experiment that compares the performance of the following estimators:

- The OLS estimator of Y on X and a constant.
- The TSLS estimator with Z_1 as an instrument for X and has only a constant as a control (included instrument).
- The TSLS estimator with $Z_1, \ldots, Z_{20}$ as instruments for X, and only a constant as a control variable.
- The jacknife IV estimator with Z_1 as an instrument for X, and only a constant as a control variable.
- The jackknife IV estimator with $Z_1, \ldots, Z_{20}$ as instruments for X, and only a constant as a control variable.

For each estimator, report the bias, median, standard deviation across simulations. Also, report the coverage rate of a 95% confidence interval. Consider sample sizes, $N = 100, 200, 400, 800$. Design an easy-to-read table to report your results. Explain your results in the context of instrumental variable regressions with many instruments.

The table below reports the results from various OLS, 2SLS, and jackknife specifications. I take 10,000 Monte Carlo draws[1] and calculate the median estimate, the bias from the true value of 1, the standard deviation of the estimates across all draws, and the coverage. To get coverage, I calculate the 95% CI using asymptotic (non-robust) standard errors on each Monte Carlo draw. I then compute the percentage of draws in which the true parameter value was in the confidence interval for each estimator.

The OLS estimator is so severely biased that coverage is 0 and the true parameter value is never in the CI. The TSLS estimators (jackknife and regular) both have the expected coverage when using 1 instrument. However, the regular TSLS estimator has bad coverage when there are many instruments. Though the jackknife estimator with multiple instruments doesn't get to a 95% coverage rate, it does significantly improve upon the regular TSLS estimator. This confirms what we covered in class–that the jackknife estimator can adjust for bias stemming from having multiple instruments with a fixed sample.

[1] I lower this value in the code sample I'm submitting so that it can run faster.

	Sample Sizes			
	N=100	N=200	N=400	N=800
OLS, No Instruments				
Median	1.5836	1.5885	1.583	1.5871
Bias	0.5799	0.5891	0.585	0.5878
SD	0.0591	0.0434	0.036	0.0235
Coverage	0	0	0	0
TSLS, 1 Instrument				
Median	1.0243	1.0199	0.9994	0.9862
Bias	0.0016	0.0029	-0.0103	-0.0149
SD	0.1655	0.1266	0.0893	0.0663
Coverage	0.93	0.91	0.95	0.92
Jackknife TSLS, 1 Instrument				
Median	0.9778	0.9949	0.9866	0.9804
Bias	-0.05	-0.0223	-0.0223	-0.0208
SD	0.1926	0.1375	0.0927	0.0674
Coverage	0.93	0.95	0.95	0.93
TSLS, Many Instruments				
Median	1.2707	1.1847	1.0918	1.036
Bias	0.2697	0.1771	0.0827	0.0385
SD	0.1029	0.0879	0.0733	0.0581
Coverage	0.33	0.46	0.73	0.85
Jackknife TSLS, Many Instruments				
Median	0.9753	1.0169	0.981	0.9798
Bias	-0.1264	-0.0255	-0.0296	-0.0191
SD	0.389	0.1626	0.1033	0.0678
Coverage	0.65	0.84	0.88	0.91

This table reports medians, SDs, and average bias from various estimators using Monte Carlo simulations of different sample sizes. For each sample size, I draw D =10,000 samples to calculate these numbers. Bias is the average bias across draws and SD is the standard deviation of the estimator across draws. To calculate coverage, I output a 95% CI for each sample and draw and store whether or not the CI contains our true mean of 1. The coverage numbers reported are the proportion of draws for which 1 was included in the asymptotic confidence interval.

Question 7 Read "Semiparametric instrumental variable estimation of treatment response models" by Alberto Abadie (2003 , Journal of Econometrics). The paper is on Canvas. The data used in Section 6 is also on Canvas.

(a) Reproduce Table 2 Note: As usual, getting standard errors exactly right can be difficult, in part because the precise meaning of "robust standard errors" is a bit vague. In particular, several small-sample modifications to the Eicker-Huber-White standard errors have been proposed in the literature, and people are usually not clear on which one they are using. These different refinements have names like HC0, HC1, HC2, etc. The best guess on what someone used is typically the Stata default, which at least today is HC1. This is also typically the form you will see in textbooks, with the degrees of freedom correction for the number of regressors. However, for older papers this is not always so clear due to changes in software over time. If you have trouble getting the standard errors, then just bootstrap!

Below I reproduce Table 2 from the Abadie paper. The estimates in columns 1-3 align with the paper, including the heteroskedasticity-robust (HC1) standard errors. The last 2 columns are both attempts at replicating the Local Average Response Function (LARF) or Least Squares Treated (LST) estimators reported in Table 2, Column 4 of the paper. In the LST (Nonpar) column, I report estimates for the semiparametric procedure with a nonparametric first stage described in Section 4.3.2 of the paper. I used the polynomial sieve estimator from the previous problem set with $K = 9$ (the tuning value that most closely matched the paper's estimates). In the LST (Par) column, I instead estimate the first stage using a logit estimator, using predicted propensity in the remaining computations. Neither of these columns exactly replicates the paper's results, and it is not clear exactly what specification the authors use (though it seems they used something nonparametric). However, the majority of the estimates from both methods fall within the 95% confidence intervals that can be calculated from the paper's reported standard errors.[2] In these final columns, I report bootstrapped standard errors run with 100 replications.

[2]Only the coefficients on family income and age from the nonparametric method fall outside the paper's 95% CI.

		Endogenous treatment			
		2SLS			
	OLS	First stage	Second Stage	LST (Nonpar)	LST (Par)
Participate in 401(k)	13,527.045 (1,809.590)		9,418.828 (2,152.081)	9,105.809 (2,446.161)	9,556.437 (1,811.293)
Constant	-23,549.003 (2,177.258)	-0.0306 (0.0087)	-23,298.738 (2,166.575)	-26,946.846 (4,224.336)	-25,396.874 (2,110.286)
Family Income	976.931 (83.335)	0.00133 (0.00014)	997.190 (83.824)	1,235.055 (118.640)	1,011.520 (79.844)
Age	-376.165 (236.888)	-0.00216 (0.00101)	-345.955 (238.009)	-574.083 (497.339)	56.473 (266.951)
Age squared	38.699 (7.663)	5e-05 (3e-05)	37.852 (7.694)	43.191 (15.488)	30.834 (8.440)
Married	-8,369.471 (1,829.238)	-0.00047 (0.00788)	-8,355.871 (1,828.980)	-9,649.525 (3,250.363)	-7,477.747 (2,438.151)
Family Size	-785.650 (410.623)	6e-05 (0.00238)	-818.962 (410.387)	-1,649.867 (790.648)	-1,056.822 (600.915)
Eligibility for 401(k)		0.688 (0.00798)			

The table above replicates Table 2 from Abadie et al. (2003). The first column presents OLS estimates. Columns (2) and (3) present the first stage and second stage estimates from a regular 2SLS. All reported standard errors in columns 1-3 are heteroskedasticity robust (HC1). Column (4) presents the results of the LARF estimator with a nonparametric first stage estimating $\tau_0(x_i)$ with a sieve polynomial of degree 9. Column (5) presents the LARF estimator with a parametric first stage that estimates $\tau_0(x_i)$ using a logit model. Standard errors in columns 4-5 are from 100 bootstrap replications.

(b) Compute a 95% Anderson-Rubin confidence interval for the coefficient on the endogenous variable (participation in a 401(k) plan). Compare this confidence interval to the one you would obtain using standard asymptotic approximation for TSLS. Are they similar or different? Explain.

I compute the 95% AR confidence interval based on the formula in the supplemental notes:

$$\text{AR}(\bar{\beta}) \equiv n\hat{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\hat{g}(\bar{\beta})$$

where $\hat{g}(\bar{\beta})$ is the OLS estimator from a regression of $Y - X'\bar{\beta}$ on Z and $\hat{\Omega}$ is the heteroskedastic-robust standard error estimator from the same regression. I use the output from the 2SLS estimates as the baseline and then create a grid to search over in order to find which values fall below the $\chi^2(k)$ statistic. I search values ± 5000 away from the 2SLS estimates with a step size of 0.5.[3] The resulting confidence interval is:

$$\text{AR CI} = [5318.918, 13510.737]$$

Compare this to the 95% CI computed from the table above:

$$\text{Asymptotic CI} = [5200.749, 13636.907]$$

The two confidence intervals are very similar, with the AR interval being slightly tighter. This makes sense since we do not have a weak instrument problem in the first stage.

[3] I changed this in the code to make it a much coarser step size because otherwise it takes a significant amount of time to run.

(c) Compute a jackknife IV estimator with bootstrapped standard error. Compare this to the TSLS results reported in Abadie's column (3). Are they similar or different? Explain.

The table below compares the TSLS (second stage reported in Column (3) of the Table 2 replication) to a jackknife TSLS with bootstrapped standard errors. The estimates and standard errors are extremely close to each other. Since we do not have a multiple instruments scenario, using the jackknife estimator does not help.

	TSLS	Jackknife TSLS
Participate in 401(k)	9,418.828	9,418.528
	(2,152.081)	(2,150.902)
Constant	-23,298.738	-23,298.720
	(2,166.575)	(2,166.679)
Family Income	997.190	997.191
	(83.824)	(83.809)
Age	-345.955	-345.952
	(238.009)	(238.032)
Age squared	37.852	37.852
	(7.694)	(7.695)
Married	-8,355.871	-8,355.870
	(1,828.980)	(1,828.986)
Family Size	-818.962	-818.965
	(410.387)	(410.393)

Column (1) reports second-stage TSLS estimates with heteroskedasticity-robust (HC1) standard errors in parentheses below. Column (2) reports second-stage jackknife TSLS estimates with bootstrapped standard errors run for 100 replications.